AFRL-AFOSR-VA-TR-2015-0233

Applications of Quantum Probability Theory to Dynamic Decision Making

Jerome Busemeyer
TRUSTEES OF INDIANA UNIVERSITY

08/13/2015
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTC
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
	Final Performance	15-07-2012 to 14-07-2015

4. TITLE AND SUBTITLE
Applications of Quantum Probability Theory to Dynamic Decision Making

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0397

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Jerome Busemeyer

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
TRUSTEES OF INDIANA UNIVERSITY
205C BRYAN HALL 105 N INIANA AVE
BLOOMINGTON, IN 47405-1106 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The broad term goal of this research program was to build a foundation for constructing probabilistic-dynamic systems from principles based on
quantum as opposed to classical probability theory. So far we have applied these principles to both traditional, one-stage decision problems studied
by decision researchers as well as dynamic Markov decision problems used in computer science and engineering. The more specific goal of the
proposed research was to develop new applications of quantum probability applied to dynamic decision situations:
(a) To develop a quantum reinforcement learning model for learning a sequence of actions in a Markov decision problem environments that is fast
learning and robust with respect to changes in the environment;
(b) To theoretically derive the convergence and speed of convergence properties of the new quantum learning algorithm for the dynamic
environments; and most importantly,
(c) To experimentally test whether the quantum reinforcement learning model provides a better account of actual human performance in Markov de

15. SUBJECT TERMS
Quantum probability, Decision models, Reinforcement learning

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Jerome Busemeyer
a. REPORT	b. ABSTRACT	c. THIS PAGE			
Unclassified	Unclassified	Unclassified	UU		19b. TELEPHONE NUMBER *(Include area code)* 812-855-4882

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

DISTRIBUTION A: Distribution approved for public release.

Statement of Objectives

Ever since Kahneman and Tversky's extremely influential research (1974, and supported by AFOSR) exposing the failures of classical probability to describe human reasoning and decision making under uncertainty, researchers gave up almost all hope to find an axiomatic foundation for understanding human judgments and decisions. Separate and disconnected heuristic explanations have been proposed using variants of classical decision theory to explain a number of paradoxical findings, such as violations of the classical probability laws of commutativity and distributivity. The paradoxical findings have resisted explanation under a common classical theoretical framework. Our past research (supported by AFOSR in the past three years) applies mathematical principles from quantum theory to cognitive and decision sciences. Our findings demonstrate that quantum theory provides a viable new direction toward the possibility of accounting for paradoxical findings from decision research using a unified and principled theoretical framework.

Research Effort

1. What is Quantum Probability Theory Applied to Decision Making Research?

Quantum probability theory (Von Neumann, 1932; Gudder, 1988; Sakurai, 1994) is unfamiliar to most cognitive, computer, and engineering scientists, so we provide a brief but general overview and a comparison with the more familiar classical probability theory (Kolmogorov, 1933). To keep it simple, we assume finite spaces although both probability theories can be extended to infinite spaces. (More details about these principles can be found in Griffiths, 2003; Gudder, 1988; Busemeyer & Bruza, 2012.)

(1) *Classical* theory begins by postulating a set called the *sample space*, Ω, which is a set of elements that contains all the events, and in the finite case this set has cardinality N. *Quantum* theory begins by postulating a *vector space* (technically, a Hilbert space), V, which contains all the events, and in the finite case this vector space has dimension N.

(2) *Classical* theory is based on the premise that an event, such as A, is a *subset* $A \subseteq \Omega$ of the sample space. *Quantum* theory is based on the premise that an event, such as A, is a *subspace* $A \subseteq V$ of the vector space. Corresponding to each subspace A is a projector, P_A, that projects points in V onto the subspace A.

(3) *Classical* theory postulates a *state* represented by a function p: $\Omega \rightarrow [0,1]$, which assigns probabilities to events in an additive manner. In other words, $p(A)$ is the probability assigned to event $A \in \Omega$, and if $A \cap B = \emptyset$, then $p(A \cup B) = p(A) + p(B)$. *Quantum* theory postulates a *state* represented by a unit length vector $\psi \in V$, which assigns probabilities to events also in an additive manner: $p(A) = ||P_A \psi||^2$ and if $A \cap B = \emptyset$ then $p(A \cup B) = p(A) + p(B)$.

(4) *Classical* theory defines a conditional state, p_A, that is a conditional probability function, as follows: If event A is observed, then $p_A(B) = p(B|A) = p(A \cap B)/p(A)$. Bayes's rule follows from this definition. *Quantum* theory defines a conditional state, ψ_A, as follows: If event A is observed, then $\psi_A = P_A \psi / \sqrt{p(A)}$, so that $p(B|A) = ||P_B P_A \psi||^2 / p(A)$.

(5) According to *classical* theory, if A,B are two events in Ω, then we can always define the *intersection* event $A \cap B = B \cap A$, and $p(A \cap B) = p(A)p(B|A) = p(B)p(A|B) = p(B \cap A)$, so the order of events does not matter. According to *quantum* theory, if A, B are two events in V, then we can define the *sequence* of events A and then B, denoted (A, B) and $p(A,B) = p(A)p(B|A) =$

$||P_B P_A \psi||^2$ and the order of the events matters. The intersection event, A∩B = B∩A, only exists in quantum theory if $P_B P_A = P_A P_B$, that is, the projectors commute, and then there is no order effect (see Griffiths, 2003, p. 53; Niestegge, 2008, p. 247). Commutativity is a key point where the two theories diverge.

2. What Is the Evidence from Our Research?

This section reviews our research that was supported by previous funds from AFOSR to accumulate evidence for the viability of applying quantum theory to human judgment and decision behavior. In particular, we focus on *interference effects*, which are violations of *the classical law of total probability*. This law holds an important role in our theories of cognition and decision because it is the foundation of Bayesian and Markov models. This law can be empirically tested by measuring the single event A alone in one condition, and measuring the joint events (A∩B), (A∩ not B) together in another condition. Violations occur when *p*(A) from the single event condition differs from *p*(A∩B) + *p*(A∩ not B). **Below we present five lines of evidence** from our previous AFOSR work on interference effects, and our quantum account of all five effects.

The first line of evidence comes from a quantum probability theory explanation for the well-known research on probability judgment errors by Tversky and Kahneman (1983). A conjunctive fallacy occurs when a person judges the probability of the conjunction of two events to be more likely than one of the constituent events. For example, the probability that a man is over 50 years old (event O) and has a heart attack (event H) is judged more likely than the probability that a man has a heart attack, even though according to the law of total probability *p*(H) = *p*(H∩O)+*p*(H∩ not O) ≥ *p*(H∩O). The disjunction fallacy occurs when a person judges the probability of the disjunction of two events to be less likely than one of the constituent events. For example, the probability that a man is over 50 or has a heart attack is judged less likely than a man is over 50. Busemeyer, Pothos, Franco, and Trueblood (2011) developed a simple quantum probability (QP) theoretical account for these puzzling findings, and we will describe the basic idea later after we present some additional lines of evidence.

Our model of the conjunction and disjunction fallacies was developed after the facts were known, and so more important tests of the model arise from new predictions. According to the QP model, if two events are incompatible, we must predict order effects when deciding about the pair of events, e.g., *p*(Ay and then Bn) ≠ p(Bn and then Ay). However, much more important than that, the QP model must predict a very special pattern of order effects, which we call the QQ equality: *p*(Ay and then Bn) + *p*(An and then By) = *p*(Bn and then Ay) + *p*(By and then An). This is an *a priori*, precise, quantitative, and parameter free prediction about the pattern of order effects, and thus the strongest test to the QP model. Recently we have shown that our QQ equality prediction was statistically supported across a wide range of 70 national field experiments that examined question order effects (Wang & Busemeyer, 2013; Wang, Solloway, Shiffrin, & Busemeyer, 2014).

The second line of evidence is based on a categorization-decision paradigm that was designed for testing *the law of total probability* (Townsend, Silva, Spencer-Smith, & Wenger, 2000). On each trial, participants are shown pictures of faces, which vary along two dimensions (face width and lip thickness). The participants are asked to categorize the faces as belonging to either a “good” guy or “bad” guy group, and/or they are asked to decide whether to take an “attack” or “withdrawal” action. The participants are provided explicit instructions about

relations between facial features, categories, and actions. A within-subjects manipulation is used to examine two conditions. In the C-then-D condition, participants categorize the face and then make an action decision; in the D-Alone condition, participants only make an action decision.

Table 1. The categorization-decision task results. (*N* = 400 across five studies)

	p(G)	p(A|G)	p(B)	p(A|B)	p_T(A)	p(A)
Good face	.78	.36	.22	.53	.39	.39
Quantum	.80	.38	.20	.62	.43	.43
Bad face	.23	.38	.77	.60	.56	.61
Quantum	.20	.37	.80	.61	.56	.62

According to the law of total probability, the probability of attack under the D-alone condition should equal the total probability of attack obtained from the C-then-D condition. However, empirical data show that they are not equal, and the difference demonstrates the interference of categorization on the decision process. The results of our first experiment using this paradigm are reported by Busemeyer, Wang, and Lambert-Mogiliansy (2009). More recently, Wang and Busemeyer (2015) reported additional five sets of experiments with study design variations to replicate and extend our initial findings, including varying number of training trials, counterbalancing face types with categories, and manipulating the probability at the trial level vs. at the block of trials level. The aggregated results (*N* = 400) are summarized in Table 1. The row labeled "good face" represents faces that came from a population (e.g., wide faces) that were associated with the good guy category, and the row labeled "bad face" represents faces that came from a population (e.g., narrow faces) that were associated with the bad guy category. The columns labeled p(G) and p(B) indicate the probability of categorizing a face as a good vs. bad guy, and p(A|G) and p(A|B) indicate the probability of attack conditioned on being categorized as a good vs. bad guy. The column labeled p_T(A) is the total probability of attack from the C-then-D condition, and p(A) is the probability of attack under the D-alone condition.

As shown in Table 1, the probability of attack under the D-alone condition substantially exceeds the total probability ($t(399) = 4.82$, $p < .001$). More dramatic is the fact that when the face came from the bad guy population, the probability of attack in the D-alone condition is even greater than that after categorizing the face as a bad guy. The interference is positive for the attack action, $p(A) > p_T(A)$ (correspondingly, negative for withdraw, $p(W) < p_T(W)$). We (Busemeyer et al., 2009; Wang & Busemeyer, 2015) developed a specific quantum model to account for these interference effects (see quantum model probabilities in Table 1). The model is summarized later after presenting another line of evidence.

The third line of evidence comes from findings of violations of a basic "rational" axiom of decision-making, called the "sure thing" principle (Savage, 1954) that states the following: If you prefer action A over B under state of the world X, and you also prefer action A over B under the complementary state of the world ~X, then you should prefer action A over B even if the state of the world is unknown. Shafir and Tversky (1992) first examined this axiom using the prisoner dilemma (PD) game. Here we briefly describe a version of the game (Croson, 1999). Eighty individuals participated in the study and each played 2 PD games. The critical manipulation was that half were required to predict what the opponent would do and then decide on an action (P-then-D), and the other half only made an action decision (D-only). The critical comparison is between the probability of defecting under the D-only condition and the total probability of defecting under the P-then-D condition. The difference demonstrates the interference effect of prediction on decision. The row labeled "Croson" in Table 2 shows the average results from the first two payoff conditions in Croson's study. In Table 2, p(d) is the probability of predicting that the opponent would defect; p(D|d) is the probability that the player defects given the opponent has been predicted to defect; p(D|c) is the probability player defects given the opponent has been predicted to cooperate; p_T(D) is the total probability to defect, and; p(D) is the probability to defect when opponent's action was not predicted in the D-only condition. As shown in Table 2, the total probability of defecting in the P-then-D condition far exceeds the probability of defecting in the D-only condition, which demonstrates the interfering effect of prediction on decisions. The interference is negative for defection, p(D) < p_T(D) (correspondingly, positive for cooperation, p(C) > p_T(C)). The earlier results by Shafir and Tversky (1992) are summarized in row S & T of Table 2. We also replicated these findings when the human player played against a computerized agent (Busemeyer, Matthews, & Wang, 2006; see rows labeled BMW). Pothos and Busemeyer (2009) developed a quantum model, as summarized below, to account for the results (see Table 2).

Table 2. Violation of the sure thing principle.

	p(d)	p(D|d)	p(c)	p(D|c)	p_T(D)	p(D)	N
Croson	.56	.67	.44	.32	.45	.30	40
S &T	.50	.97	.50	.84	.91	.63	80
BMW1	.50	.92	.50	.84	.88	.65	88
BMW 2	.50	.88	.50	.73	.81	.65	410
Quantum	.50	.82	.50	.72	.77	.65	

A brief description of the quantum theoretical account. All three lines of evidence discussed above (conjunction/disjunction judgment, categorization-decision process, and prisoner dilemma tasks) showed violations of the law of total probability of the classical theory and interference effects. Quantum theory provides a natural account for the findings. In all three experimental paradigms, the decision maker makes an inference and then a decision. During the first stage, the decision maker is placed into one of three inference states: (1) a state ψ_1 in which one type of inference is made (e.g., the man is young, the face is a good guy, the opponent will cooperate); (2) a state ψ_2 in which the other type of inference is made (e.g., the man is old, the face is a bad guy, the opponent will defect); or (3) a *superposition* state $\psi_U = (\sqrt{a}\cdot\psi_1 + \sqrt{b}\cdot\psi_2)$ in which the decision maker remains indefinite or uncertain about the inference (e.g., the man's age, the category of a face, the disposition of an opponent), such as in the decision-alone

conditions in the experiments. Then the decision maker is asked to take a decision (e.g., decide whether or not the man will have a heart attack, decide whether or not to attack, decide whether or not to defect). If P_A represents the projector matrix for taking an action A, then the probability of taking action A from the first inference state ψ_1 equals $p(\text{A}|\text{state 1}) = ||P_A \cdot \psi_1||^2$; the probability of taking action A from the second inference state ψ_2 equals $p(\text{A}|\text{state 2}) = ||P_A \cdot \psi_2||^2$; and for the superposition state, we have $p(\text{A}|\text{superposed}) = ||P_A \cdot \psi_U||^2 = ||P_A \cdot (\sqrt{a} \cdot \psi_1 + \sqrt{b} \cdot \psi_2)||^2 = a \cdot ||P_A \cdot \psi_1||^2 + b \cdot ||P_A \cdot \psi_2||^2 + Int$, where *Int* represents the cross-product terms produced by squared length of the sum. Thus, the probability of taking the action from the uncertain state is the weighted average of the two known states (corresponding to the classical "total probability") plus interference. The interference term *Int* can be positive or negative, which is used to account for the violations of the law of total probability. Of course the critical part of the model is to **derive the interference term from basic principles**. This is exactly what was done in all three lines of research by deriving the interference term from a dynamic quantum model based on the Schrödinger equation. To further test the quantum model, **a stronger quantitative test** was conducted in the application below.

The fourth line of evidence for interference effects comes from research we conducted on a phenomenon called *dynamic inconsistency* (Barkan & Busemeyer, 1999, 2003). Most complex decisions involve multiple stages that require planning for the future across sequences of actions and events. Optimal strategies use backward induction algorithms that require planning from the last stage and working backwards to the current stage. Dynamic consistency requires that the planned actions are actually carried out once those decisions are realized. Barkan and Busemeyer (2003) investigated dynamic consistency by using a modification of a two-stage gambling paradigm originally used by Tversky and Shafir (1992). A total of 100 people participated in the experiment. Each person played 17 different gambles, and each gamble was played twice. The first play was obligatory, but the player was given a choice whether or not to play the gamble again on the second round. For each gamble, the player made two choices: a planned choice contingent on winning or losing the first stage, and a final choice after actually playing and experiencing the outcome of the first stage. The planned and the final decisions were made equally valuable because the experimenter randomly selected either the planned action or the final action to determine the final monetary payoff. A dynamic inconsistency effect occurred—people changed systematically away from their plans on the final decision. Actually winning the first stage decreased the probability of playing the gamble again at the second stage compared to the plan, while actually losing the first stage increased the probability compared to the plan. Once again, these effects, called dynamic inconsistency, were inconsistent with the law of total probability.

Quantitative model comparisons. To explain the dynamic inconsistency effects, Barkan and Busemeyer (2003) used a reference point change model based on prospect theory (originally proposed by Tversky and Shafir, 1992, for this two stage game paradigm). However, the quantum model developed by Pothos and Busemeyer (2009) for the prisoner dilemma game can also account for these results. Naturally the question is: Which model is better? To answer this, Busemeyer, Wang, and Shiffrin (2014) completed a rigorous quantitative comparison of these two competing models using the data from Barkan and Busemeyer (2003). Both models use only three parameters (a risk aversion parameter, a loss aversion parameter, and a parameter related to the choice probability function) to predict 34 data points (plan vs. final choices for 17 payoff

conditions). First, each model was fit to the means, and R^2 was used to compare fits. The R^2 for the quantum model (.82) substantially exceeded the R^2 for the reference point change model (.77). Second, a Bayes factor, BF = p(quantum model | data) / p(reference point model | data) was computed for each participant, where p(model | data) equals the expected likelihood for a model based on the sequence of 66 planned and final choices made by each participant to the 17 gambles. The Bayes factor was computed using both uniform and normal priors on the parameters. In both cases, the Bayes factor strongly supported the quantum model. For the uniform prior, the total (across participants) log Bayes factor equaled 74.5 and over 90% of the participants produced positive log Bayes factors with this prior; for the normal prior, the total log Bayes factor equaled 83.05 and over 93% of the participants produced positive log Bayes factors with this prior (Busemeyer et al., 2014).

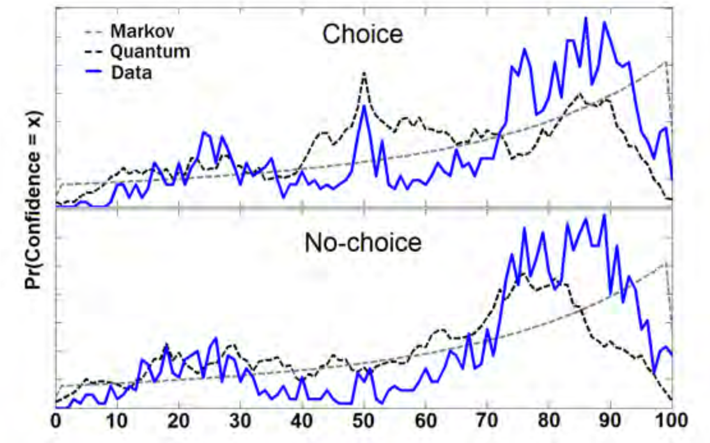
The fifth line of evidence extended our testing of (the predicted violations of) the law of total probability with dynamic decision problems. The new evidence was recently obtained from research on signal detection type tasks in which a decision maker must decide whether a target is present or absent based on noisy and uncertain information (e.g., to decide whether an enemy is located at a position based on a poor and fuzzy image). Human performance (accuracy, decision time, and confidence) observed with signal detection tasks has traditionally been modeled using Markov type of random walk models of decision-making (e.g., see Busemeyer & Townsend, 1993; Pleskac & Busemeyer, 2010). The basic idea is that the decision maker accumulates evidence for each hypothesis until the accumulated evidence reaches a threshold. The first hypothesis to reach the threshold is chosen and the time to reach the threshold determines the decision time, and the difference in evidence soon after the decision determines the confidence. Alternatively, Busemeyer, Wang, and Townsend (2006) developed a quantum random walk model for signal detection, which assumes that a person's evidence state is represented by a wave function spread over levels of evidence. The Markov model evolves probabilities over time according to the Kolmogorov forward equation, and the quantum model evolves amplitudes over time according to the Schrödinger equation.

Busemeyer and Bruza (2012, ch. 8) derived a key prediction that provides a critical method to empirically distinguish and test the two theories. The experiment consists of two conditions: In the choice-confidence condition, the person makes a choice (signal present or absent) at time t_1 and then rates confidence at time t_2; in the confidence-alone condition, the person only provides a confidence rating at time t_2. For both conditions, the focus is on the marginal distribution of confidence ratings that are obtained at time t_2. Confidence is defined as the probability that a signal is present on a scale ranging from 0 = the target is not present, to 50 = undecided, to 100 = the target is present. The Markov model obeys *the Chapman-Kolmogorov equation*, which is a dynamic form of the law of total probability and predicts *no difference* between the two conditions. The quantum model predicts that an interference effect is produced by decision on the confidence rating that makes the confidence distributions differ between the two conditions.

Kvam, Pleskac, Yu, and Busemeyer (2015) empirically tested this prediction and obtained strong support for the predicted interference effect. Figure 1 shows the density of a participant's confidence responses (in blue) over each confidence level, scaled such that 0 is complete certainty in target absent and 100 is complete certainty target present. Model predictions for the quantum random walk (black dashed) and Markov random walk (gray dashed) are also shown, which are based on the maximum likelihood estimates for each model.

Figure 1 clearly illustrates the interference, and the interference was statistically significant for seven out of nine participants (each participant contributing over 2500 trials).

Figure 1. Interference effects of choice on subsequent confidence.



3. Existing Engineering Applications of Quantum Decision Theory to the Predator – Prey Dynamic Game

The applications of quantum decision theory described so far have been restricted to fairly simple, basic decision situations. In our previous work supported by AFOSR, we have also examined applications of quantum decision theory to more complex dynamic decision problems within the class of Markov decision problems (MDP's) (Fakhari, Rajagopal, Balakrishnan, & Busemeyer, 2013). This class of problems includes situations such as predator-prey target tracking and goal seeking tasks, which are relevant to Air Force applications. In particular, we developed a new quantum reinforcement learning algorithm for MDP's. The quantum reinforcement-learning algorithm does not require a quantum computer, and can be directly used to learn to perform practical sequential decision-making tasks. Our research, summarized below, indicates that the proposed quantum reinforcement learning algorithm is more robust for learning optimal strategies in complex dynamic decision environments than traditional models.

The quantum reinforcement learning algorithm. It uses the same Q-learning algorithm to estimate values of actions as used in traditional reinforcement learning models (Sutton & Barto, 1998). The key difference is concerned with the probabilistic rules to select actions. Unlike traditional models that use, for example, the epsilon greedy algorithm, or the soft max rule for action selection, the quantum model uses quantum probability rules for selecting actions. The idea of using a quantum rule for action selection was first proposed and tested by Dong et al. (2008). We have, however, made major modifications to substantially improve Dong's original algorithm. The basic idea is that the current environmental state puts the agent in a superposition state over the set of possible actions. The superposition state is a vector in an m dimensional space spanned by m orthonormal basis vectors denoted $|a_k\rangle$, $k=1,...m$ and each basis vector corresponds to one of the actions. If the current environmental state is e_j, then the superposition state over actions is $|\psi_j\rangle=\sum_{k=1,m} \psi_{jk}\cdot|a_k\rangle$, with two constraints on the amplitudes: $\psi_{jk}=0$ for any action that is not available from state e_j, and given the previous constraint, we also require $|\psi_j\rangle$ to remain unit length. Then the probability of taking action a_k from state e_j equals $|\psi_{jk}|^2$. The key new idea is the updating rule for modifying the amplitudes ψ_{jk} that experience

rewards. Hereafter, the $m\times 1$ column matrix ψ will refer to the amplitudes for m actions and each action is assumed to be a potential choice.

Amplitude amplification. The amplitude amplification algorithm is an extension of Grover (1997)'s quantum information search algorithm (Hoyer, 2000). The algorithm begins with any arbitrary initial amplitude distribution represented by the $m\times 1$ column matrix ψ_0, but it is common to start with $\psi = (1/(\sqrt{m}))$ for m actions. Define ψ_t as the $m\times 1$ matrix of amplitudes after experiencing t trials of training. Suppose action a_j was chosen on the last trial t. The amplitude for action a_j is amplified or attenuated in proportion to reward $[r(t)+\gamma\cdot\max_l Q(e,a_l,t)]$ experienced by taking that action, where $Q(e,a_l,t)$ is the value of an action learned by a temporal difference Q learning algorithm. The amplification computed as follows. Define A_k as an $m\times 1$ matrix with zeros in every row except the row k corresponding to action a_k, which is set equal to one. This is essentially the coordinates corresponding to the basis vector $|a_k\rangle$. Next define two matrices

$$Q_1 = I - (1 - \exp\{i\varphi_1\})\cdot(A_k\cdot A_k^{\dagger}), \text{ and } Q_2 = (1 - \exp\{i\varphi_2\})\cdot(\psi_t\cdot\psi_t^{\dagger}) - I, (2)$$

where φ_1, φ_2 are two learning parameters that control the amount of amplification or attenuation. Then the new amplitude distribution is formed by $\psi_{t+1} = (Q_2\cdot Q_1)^L\cdot\psi_t$, where the matrix power L indicates the integer number of applications of the update used on a single trial. The new idea is to relate the parameters (L, φ_1, φ_2) to the Q value of the selected action. Dong et al. (2008) proposed to map Q values into the parameter L, which is an integer number of amplifications. However, this becomes very problematic for small numbers of actions. Also this method only amplifies and never attenuates the amplitude assigned to an action. Instead, our new model fixes L at one, and we map normalized values of Q from the Q-learning algorithm into the two phases φ_1, φ_2 to amplify rewarded actions and to attenuate actions that are punished. The *key idea* for robustness is that for a given number of actions, N, the mapping from the Q values of the Q-learning model to the parameters φ_1 and φ_2 can be determined *a priori* to provide *robust learning*. Unlike the epsilon greedy and softmax rules, the quantum parameters do not need to be adjusted post hoc for each variation in the environment.

Evaluating quantum algorithm. To evaluate our quantum algorithm practically, we conducted computer simulations within a large grid world, using a prey-predator game involving two competing predators and one randomly moving prey. The predators are given information about the distance from the prey in each direction on each step. One predator was based on the traditional soft max probabilistic choice rule, and the other was based on our new quantum probabilistic rule. (We also compared results with the epsilon greedy choice rule, but this did not perform as well as the soft max rule, and so we focus on the latter.) The aim of the task is to find a policy that will let the predator find the prey with minimum punishment. Fakhari et al. (2013) conducted extensive simulations varying the size of the grid world and the number of actions. The main results are summarized in Table 3, which shows the number of times each agent captured the prey when both agents were competing to catch the same prey. At the early stage of training on the task (learning Q values), the soft max algorithm caught more prey than our quantum algorithm; however, at intermediate and later stages, the quantum algorithm strongly outperformed the soft max rule.

Table 3. Winning statistics of modified QRL and softmax agents.

Test cases	NM-QRL alone winning	Softmax alone winning	Both agents winning	Average no. of steps	
				NM-QRL	Softmax
Initial stage	14722	31795	3482	32.76	41.23
Intermediate stage	28463	11129	10407	18.27	19.73
Final stage	33511	15494	994	10.53	10.71

6. New Engineering Applications of Quantum Decision Theory to Target Assignments

The concept of designing a group of intelligent systems with coordinating action capabilities is called "cooperative control" (Arslan, Marden, & Shamma, 2007; Olfati-Saber, 2006). We have considered (Rajagopol, Balakrishnan, Busemeyer, 2015) an example problem where a group of mobile agents with motion uncertainty should dynamically assign themselves to unique target points. This problem can be viewed as a combinatorial optimization problem. The target/task assignment combinatorial problems are non-deterministic polynomial time complete (Murphey, 2000). Traditional approaches use heuristic methods to quickly obtain sub-optimal assignment profiles. However, these approaches require a centralized decision-making framework wherein individual agents have access to information about all other agents. The computational complexity of centralized decision-making process increases with the increase in number of agents.

To combat the limitations of centralized approaches, decentralized approaches based on multi-player game theory (Fudenberg & Tirole, 1991; Basar & Olsder, 1999) are recommended for multi-agent problems (Arslan et al., 2007). Then, the optimal assignment profile is equated to the pure Nash equilibrium of the multiplayer-game i.e. each agent chooses the best assignment taking into consideration the assignment of other agents and no agent will benefit by unilaterally changing its assignment. A general theme in multi-player learning algorithms is that each agent should empirically model the response of other agents. Then, the agent can use the empirical model to choose its best response that will maximize its expected utility. When the target assignment problem is formulated as a multi-player game, multi-player learning algorithms like Fictitious-play (Fudenberg & Levine, 1998) can be used to design the negotiation mechanism. However, for large numbers of agents, the maximization of expected utility in real time can be a time consuming process and these learning algorithms can be computationally expensive. Quantum decision theory provides a natural computational framework to implement the proposed approach. This idea is demonstrated using an example coordination problem.

Consider two planar robots moving in an uncertain environment to reach two different goal states. Let $x_a \in \mathbb{R}^2$ with $a = 1,2$ represent the current position information of the robots and μ_s with $s = 1,2$ represent the goal states. It is assumed that the robots are fully capable of reaching any of the goal states. At time t, $p(\mu_s / x_a, t)$ represent the probability that the agent 'a' will choose target μ_s. Here, the probability distribution depends upon the robot's utility function. The global objective is that the agents should reach unique targets within some time frame. The action choices for the robots are choose target '1' or choose target '2'.

To formulate the above problem in terms of quantum decision theory, standard notations a decision-state of any robot is represented by the Ket vector notation $|.\rangle$. The rational choice for each robot would be, at each instant to choose a target corresponding to the maximal of the

probability distribution $p(\mu_s / x_a, t)$. However, there are two problems associated with such an action selection mechanism. Both the agents might select the same target state as the goal state. Another issue will be that since the state x_a evolves in an uncertain environment the probability distribution $p(\mu_s / x_a, t)$ might not be stationary. Our desired composite state i.e. the decision-state that will reflect the global objective is given by

$$\left|\psi_{12}^{d}\right\rangle = \tau_1 \left|\mu_1, \mu_2\right\rangle + \tau_2 \left|\mu_2, \mu_1\right\rangle \quad (1)$$

The above form ensures that if robot ‘1’ chooses μ_1, then robot ‘2’ will definitely choose μ_2 and vice versa irrespective of their rational choices. In quantum mechanics, Eq. (1) is called the Einstein-Podolsky-Rosen state and is a famous example of entangled state. For implementing the above idea, assume that each agent has an independent target selection mechanism. However, the target selection mechanism should result in a composite representation consistent with Eq. (1). For achieving that, each agent models the influence of other robot’s behavior on its action choices by an entanglement factor. Let $\gamma_1^{\mu_1}(t)$ and $\gamma_1^{\mu_2}(t)$ represent the entanglement factors as perceived by robot ‘1’ for action choices $|\mu_1\rangle$ and $|\mu_2\rangle$ respectively. The new, entangled composite decision-state representation for agent 1 is given by,

$$\left|\psi_{12}^{1}\right\rangle = i \sin\left(\frac{\gamma_1^{\mu_1}(t)}{2}\right)\left|\mu_1, \mu_2\right\rangle + \sin\left(\frac{\gamma_1^{\mu_2}(t)}{2}\right)\left|\mu_2, \mu_1\right\rangle \quad (2)$$

The updating of entanglement factors is proposed below:

i) Initialize the entanglement factors: $\gamma_1^{\mu_1}(t_0) = \gamma_1^{\mu_1}(t_0) = \gamma_1^{\mu_1}(t_0) = \gamma_1^{\mu_1}(t_0) = \pi / 2$.

At every time instant t

ii) Based on $p(./ x_1 : x_2, t)$ assign each robot a unique target. Hence, the robot should communicate among them the $p(./ x_a, t)$ values at every time instant.

iii) Based on the assigned target increase/decrease the corresponding entanglement factors. For example, if at time t robot ‘1’ is assigned to goal state μ_1 and robot ‘2’ is assigned to goal state μ_2 then the respective entanglement factors are updated in the following way:

$$\begin{aligned} \dot{\gamma}_1^{\mu_1}(t) &\propto p(./ x_1, t)^T p(./ x_2, t), \dot{\gamma}_1^{\mu_2}(t) = -\dot{\gamma}_1^{\mu_1}(t) \\ \dot{\gamma}_2^{\mu_2}(t) &\propto p(./ x_1, t)^T p(./ x_2, t), \dot{\gamma}_2^{\mu_1}(t) = -\dot{\gamma}_1^{\mu_2}(t) \end{aligned} \quad (3)$$

The above example demonstrates how quantum decision theory can be employed for multi-agent task assignment problems. For demonstration the proposed approach was compared the potential game theory approach described in Arslan et al. (2007). The simulation was performed for a scenario where are three robots and three target points. The objective is each robot should reach a unique target point. It was assumed that robots have motion uncertainty. For potential game theory approach, a utility function similar to that defined in Arslan et al. (2007) was used. 100 sample cases were run for comparison. Using the quantum decision theory method, we observed that the robots reach unique targets. Also, it was observed that the entanglement method performed way better than the potential game theory approach. The results indicated that with potential game theory approach, for at least 35 cases robots were quite far from the target points at the final time. Note that Eq. (3) is just one way of updating the entanglement factor. Our research work has concentrated on optimally updating the entanglement factor keeping in line with the objective of “minimization of deviation from rational decisions.”

Publications supported by the Grant

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences, 36*, 255-274. (Target Article).

Fakhari, P., Rajagopal, K., Balakrishnan, S. N., & Busemeyer, J. R (2013). Quantum inspired reinforcement learning in changing environments. *New Mathematics and Natural Computation: Special Issue on Engineering of the Mind, Cognitive Science and Robotics, 9* (3), 273-294.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review*, *120* (3), 679-696.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science, 5*(4), 672-688.

Pothos, E. M., Shiffrin, R. M, & Busemeyer, J. R. (2014). The dynamics of decision making when the probabilities are vaguely specified. *Journal of Mathematical Psychology, 59,* 6-17.

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Science, 111*(26), 9431-9436. doi: 10.1073/pnas.1407756111

Busemeyer, J. R., & Wang, Z. (2015). What is quantum cognition and how is it applied to psychology? *Current Directions in Psychological Science*, *24* (3), 163-169.

White, L., Pothos, E. M., & Busemeyer, J. R. (2015) Insights from quantum cognitive models for organizational decision making. *Journal of Applied Research in Memory and Cognition*.

White, L.C., Pothos, E.M., & Busemeyer, J.R. (2014). Sometimes it does hurt to ask: The constructive role of articulating impressions. *Cognition*, 133(1), 48-64.

Busemeyer, J. R., Wang, Z. Pothos, E. M., & Trueblood, J. T. (2015). The conjunction fallacy, confirmation, and quantum theory: Comment on Tentori, Crupi, & Russo (2013). *Journal of Experimental Psychology: General*, *144*(1), 236-243. doi:http://dx.doi.org/10.1037/xge0000035

Busemeyer, J. R., Wang, Z. Khrennikov, A., & Basieva, I. (2014). Applying quantum principles to psychology. *Physica Scripta*, T163, 014007

Khrennikov, A., Basieva, I., Dzhafarov, E., & Busemeyer, J. R. (2014). Quantum models for psychological measurements: An unsolved problem. *PLoS*, 9 (10) e110909

Busemeyer, J. R., Wang, Z., & Shiffrin, R. S. (2015). Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency. *Decision*, 2, 1-12.

Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (in press) Interference Effects of Choice on Confidence. *Proceedings of the National Academy of Science.*

Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015) Quantum cognition: A new theoretical approach to psychology. *Trends in Cognitive Science*, *19* (7), 383-393.

Rajagopal, K., Balakrishnana, S.N., Busemeyer, J.R. (2015) Neural network based solutions for stochastic optimal control using path integrals. Under Revision for IEEE Transactions on Neural Networks and Learning Systems.

Rajagopal, K., Balakrishnan, S. N., Busemeyer, J. R. (2015) Dynamic target assignment for multi-agent systems in stochastic environment using quantum inspired negotiation models. Research Report.

1.

1. Report Type

F na Report

Primary Contact E-mail
Contact email if there is a problem with the report.

jbusemey@ nd ana.edu

Primary Contact Phone Number
Contact phone number if there is a problem with the report

8128554882

Organization / Institution name

Ind ana Un vers ty

Grant/Contract Title
The full title of the funded effort.

App cat ons of Quantum Probab ty Theory to Dynam c Dec s on Mak ng

Grant/Contract Number
AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-12-1-0397

Principal Investigator Name
The full name of the principal investigator on the grant or contract.

Jerome R. Busemeyer

Program Manager
The AFOSR Program Manager currently assigned to the award

James H Lawton

Reporting Period Start Date

07/15/2012

Reporting Period End Date

07/14/2015

Abstract

The broad term goa of th s research program was to bu d a foundat on for construct ng probab st c-dynam c systems from pr nc p es based on quantum as opposed to c ass ca probab ty theory. So far we have app ed these pr nc p es to both trad t ona , one-stage dec s on prob ems stud ed by dec s on researchers as we as dynam c Markov dec s on prob ems used n computer sc ence and eng neer ng. The more spec f c goa of the proposed research was to deve op new app cat ons of quantum probab ty app ed to dynam c dec s on s tuat ons:
(a) To deve op a quantum re nforcement earn ng mode for earn ng a sequence of act ons n a Markov dec s on prob em env ronments that s fast earn ng and robust w th respect to changes n the env ronment;
(b) To theoret ca y der ve the convergence and speed of convergence propert es of the new quantum earn ng a gor thm for the dynam c env ronments; and most mportant y,
(c) To exper menta y test whether the quantum re nforcement earn ng mode prov des a better account of actua human performance n Markov dec s on prob ems as compared to more trad t ona earn ng mode s.

Distribution Statement
This is block 12 on the SF298 form.

D str but on A - Approved for Pub c Re ease

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

AFD-070820-035.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

F na Report.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probab   ty prov de a new d rect on for cogn t ve mode ng? Behav ora and Bra n Sc ences, 36, 255-274. (Target Art c e).

Fakhar , P., Rajagopa , K., Ba akr shnan, S. N., & Busemeyer, J. R (2013). Quantum nsp red re nforcement earn ng n chang ng env ronments. New Mathemat cs and Natura Computat on: Spec a Issue on Eng neer ng of the M nd, Cogn t ve Sc ence and Robot cs, 9 (3), 273-294.

Pothos, E. M., Busemeyer, J. R., & Trueb ood, J. S. (2013). A quantum geometr c mode of s m ar ty. Psycho og ca Rev ew, 120 (3), 679-696.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potent a of us ng quantum theory to bu d mode s of cogn t on. Top cs n Cogn t ve Sc ence, 5(4), 672-688.

Pothos, E. M., Sh ffr n, R. M, & Busemeyer, J. R. (2014). The dynam cs of dec s on mak ng when the probab   t es are vague y spec f ed. Journa of Mathemat ca Psycho ogy, 59, 6-17.

Wang, Z., So oway, T., Sh ffr n, R. M., & Busemeyer, J. R. (2014). Context effects produced by quest on orders revea quantum nature of human judgments. Proceed ngs of the Nat ona Academy of Sc ence, 111(26), 9431-9436. do : 10.1073/pnas.1407756111

Busemeyer, J. R., & Wang, Z. (2015). What s quantum cogn t on and how s t app ed to psycho ogy? Current D rect ons n Psycho og ca Sc ence, 24 (3), 163-169.

Wh te, L., Pothos, E. M., & Busemeyer, J. R. (2015) Ins ghts from quantum cogn t ve mode s for organ zat ona dec s on mak ng. Journa of App ed Research n Memory and Cogn t on.

Wh te, L.C., Pothos, E.M., & Busemeyer, J.R. (2014). Somet mes t does hurt to ask: The construct ve ro e of art cu at ng mpress ons. Cogn t on, 133(1), 48-64.

Busemeyer, J. R., Wang, Z. Pothos, E. M., & Trueb ood, J. T. (2015). The conjunct on fa acy, conf rmat on, and quantum theory: Comment on Tentor , Crup , & Russo (2013). Journa of Exper menta Psycho ogy: Genera , 144(1), 236-243. do :http://dx.do .org/10.1037/xge0000035

Busemeyer, J. R., Wang, Z. Khrenn kov, A., & Bas eva, I. (2014). App y ng quantum pr nc p es to psycho ogy. Phys ca Scr pta, T163, 014007

Khrenn kov, A., Bas eva, I., Dzhafarov, E., & Busemeyer, J. R. (2014). Quantum mode s for psycho og ca measurements: An unso ved prob em. PLoS, 9 (10) e110909

Busemeyer, J. R., Wang, Z., & Sh ffr n, R. S. (2015). Bayes an mode compar son favors quantum over standard dec s on theory account of dynam c ncons stency. Dec s on, 2, 1-12.

Kvam, P. D., P eskac, T. J., Yu, S., & Busemeyer, J. R. (n press) Interference Effects of Cho ce on Conf dence. Proceed ngs of the Nat ona Academy of Sc ence.

Bruza, P. D., Wang, Z., & Busemeyer, J. R. (2015) Quantum cogn t on: A new theoret ca approach to psycho ogy. Trends n Cogn t ve Sc ence, 19 (7), 383-393.

Rajagopa , K., Ba akr shnana, S.N., Busemeyer, J.R. (2015) Neura network based so ut ons for stochast c opt ma contro us ng path ntegra s. Under Rev s on for IEEE Transact ons on Neura Networks and Learn ng Systems.

Rajagopa , K., Ba akr shnan, S. N., Busemeyer, J. R. (2015) Dynam c target ass gnment for mu t -agent systems n stochast c env ronment us ng quantum nsp red negot at on mode s. Research Report

Changes in research objectives (if any):

none

Change in AFOSR Program Manager, if any:

none

Extensions granted or milestones slipped, if any:

none

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Start ng FY	FY+1	FY+2
Sa ary			
Equ pment/Fac t es			
Supp es			
Tota			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Aug 10, 2015 08:28:37 Success: Ema Sent to: jbusemey@ nd ana.edu